letters to nature

1. Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
2. Briscoe, J. & Ericson, J. The specification of neuronal identity by graded Sonic Hedgehog signalling. *Semin. Cell Dev. Biol.* **10**, 353–362 (1999).
3. Strigini, M. & Cohen, S. M. A Hedgehog activity gradient contributes to AP axial patterning of the *Drosophila* wing. *Development* **124**, 4697–4705 (1997).
4. Vervoort, M., Crozatier, M., Valle, D. & Vincent, A. The COE transcription factor Collier is a mediator of short-range Hedgehog-induced patterning of the *Drosophila* wing. *Curr. Biol.* **9**, 632–639 (1999).
5. Jiang, J. & Struhl, G. Protein kinase A and Hedgehog signalling in *Drosophila* limb development. *Cell* **80**, 563–572 (1995).
6. Li, W., Ohlmeyer, J. T., Lane, M. E. & Kalderon, D. Function of protein kinase A in hedgehog signal transduction and *Drosophila* imaginal disc development. *Cell* **80**, 553–562 (1995).
7. Pan, D. & Rubin, G. M. cAMP-dependent protein kinase and *hedgehog* act antagonistically in regulating *decapentaplegic* transcription in *Drosophila* imaginal discs. *Cell* **80**, 543–552 (1995).
8. Lepage, T., Cohen, S. M., Diaz-Benjumea, F. J. & Parkhurst, S. M. Signal transduction by cAMP-dependent protein kinase A in *Drosophila* limb patterning. *Nature* **373**, 711–715 (1995).
9. Wang, G., Wang, B. & Jiang, J. Protein kinase A antagonizes Hedgehog signaling by regulating both the activator and repressor forms of Cubitus interruptus. *Genes Dev.* **13**, 2828–2837 (1999).
10. Price, M. A. & Kalderon, D. Proteolysis of cubitus interruptus in *Drosophila* requires phosphorylation by protein kinase A. *Development* **126**, 4331–4339 (1999).
11. Jia, J. *et al.* Shaggy/GSK3 antagonizes Hedgehog signalling by regulating Cubitus interruptus. *Nature* **416**, 548–552 (2002).
12. Price, M. A. & Kalderon, D. Proteolysis of the Hedgehog signaling effector Cubitus interruptus requires phosphorylation by Glycogen Synthase Kinase 3 and Casein Kinase 1. *Cell* **108**, 823–835 (2002).
13. Jiang, J. Degrading Ci: who is Cul-pable? *Genes Dev.* **16**, 2315–2321 (2002).
14. Ohlmeyer, J. T. & Kalderon, D. Hedgehog stimulates maturation of Cubitus interruptus into a labile transcriptional activator. *Nature* **396**, 749–753 (1998).
15. Denef, N., Neubuser, D., Perez, L. & Cohen, S. M. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell* **102**, 521–531 (2000).
16. Ohlmeyer, J. T. & Kalderon, D. Dual pathways for induction of wingless expression by protein kinase A and Hedgehog in *Drosophila* embryos. *Genes Dev.* **11**, 2250–2258 (1997).
17. Kalderon, D. & Rubin, G. M. Isolation and characterization of *Drosophila* cAMP-dependent protein kinase genes. *Genes Dev.* **2**, 1539–1556 (1988).
18. Alcedo, J., Ayzenzon, M., von Ohlen, T., Noll, M. & Hooper, J. E. The *Drosophila smoothened* gene encodes a seven-pass membrane protein, a putative receptor for the Hedgehog signal. *Cell* **86**, 221–232 (1996).
19. van-den-Heuval, M. & Ingham, P. W. *smoothened* encodes a receptor-like serpentine protein required for *hedgehog* signalling. *Nature* **382**, 547–551 (1996).
20. Kemp, B. E. & Pearson, R. B. Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15**, 342–346 (1990).
21. Umphress, J. L., Tuazon, P. T., Chen, C. J. & Traugh, J. A. Determinants on simian virus 40 large T antigen are important for recognition and phosphorylation by casein kinase I. *Eur. J. Biochem.* **203**, 239–243 (1992).
22. Chijiwa, T. *et al.* Inhibition of forskolin-induced neurite outgrowth and protein phosphorylation by a newly synthesized selective inhibitor of cyclic AMP-dependent protein kinase, *N*-[2-(*p*-bromocinnamylamino)ethyl]-5-isoquinolinesulfonamide (H-89), of PC12D pheochromocytoma cells. *J. Biol. Chem.* **265**, 5267–5272 (1990).
23. Chijiwa, T., Hagiwara, M. & Hidaka, H. A newly synthesized selective casein kinase I inhibitor, *N*-(2-aminoethyl)-5-chloroisoquinoline-8-sulfonamide, and affinity purification of casein kinase I from bovine testis. *J. Biol. Chem.* **264**, 4924–4927 (1989).
24. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
25. Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nature Biotechnol.* **18**, 896–898 (2000).
26. Zhu, A. J., Zheng, L., Suyama, K. & Scott, M. P. Altered localization of *Drosophila* Smoothened protein activates Hedgehog signal transduction. *Genes Dev.* **17**, 1240–1252 (2003).
27. Calleja, M., Moreno, E., Pelaz, S. & Morata, G. Visualization of gene expression in living adult *Drosophila*. *Science* **274**, 252–255 (1996).
28. Jia, J., Tong, C. & Jiang, J. Smoothened transduces Hedgehog signal by physically interacting with Costal2/Fused complex through its carboxyl-terminal tail. *Genes Dev.* **17**, 2709–2720 (2003).
29. Motzny, C. K. & Holmgren, R. The *Drosophila* cubitus interruptus protein and its role in the *wingless* and *hedgehog* signal transduction pathways. *Mech. Dev.* **52**, 137–150 (1995).
30. Lum, L. *et al.* Hedgehog signal transduction via Smoothened association with a cytoplasmic complex scaffolded by the atypical kinesin, Costal-2. *Mol. Cell* **12**, 1261–1274 (2003).

Accurate multiplex gene synthesis from programmable DNA microchips

Jingdong Tian[1], **Hui Gong**[1], **Nijing Sheng**[2], **Xiaochuan Zhou**[3], **Erdogan Gulari**[4], **Xiaolian Gao**[2] & **George Church**[1]

[1]*Harvard Medical School, 77 Ave Louis Pasteur, Boston, Massachusetts 02115, USA*
[2]*Department of Chemistry, University of Houston, Houston, Texas 77004-5003, USA*
[3]*Atactic Technologies Inc., 2575 West Bellfort, Suite 270, Houston, Texas 77054, USA*
[4]*Chemical Engineering Department, University of Michigan, Ann Arbor, Michigan 48109-2136, USA*

Testing the many hypotheses from genomics and systems biology experiments demands accurate and cost-effective gene and genome synthesis. Here we describe a microchip-based technology for multiplex gene synthesis. Pools of thousands of 'construction' oligonucleotides and tagged complementary 'selection' oligonucleotides are synthesized on photo-programmable microfluidic chips[1], released, amplified and selected by hybridization to reduce synthesis errors ninefold. A one-step polymerase assembly multiplexing reaction assembles these into multiple genes. This technology enabled us to synthesize all 21 genes that encode the proteins of the *Escherichia coli* 30S ribosomal subunit, and to optimize their translation efficiency *in vitro* through alteration of codon bias. This is a significant step towards the synthesis of ribosomes *in vitro* and should have utility for synthetic biology in general.

The advance of large-scale biochemical analyses such as sequencing, microarrays and proteomics has generated vast amounts of data, which computational biologists have leveraged into a large number of hypotheses. To test these hypotheses we must first overcome the bottleneck in constructing new genetic elements, genetic pathways and engineered cells. To optimize complex biological processes using darwinian selection, we need to direct the finite diversity available in combinatorial oligonucleotide synthesis (about 25 randomized base pairs (bp) or equivalents) thoughtfully through large stretches (at the megabase level) of DNA sequence. These represent great challenges and potential payoffs for the emerging field of synthetic biology.

In principle, we can create a useful variety of molecules, cellular and cell-free systems given a sufficient supply of custom genes and genomes. However, current methods for generating even basic oligonucleotides are expensive (US$0.11 per nucleotide) and have very high levels of errors (deletions at a rate of 1 in 100 bases and mismatches and insertions at about 1 in 400 bases). As a result, gene or genome synthesis from oligonucleotides is both expensive and prone to error. Correcting errors by clone sequencing and mutagenesis methods further increases the amount of labour and total cost (to at least US$2 per base pair). In principle, the cost of oligonucleotide synthesis can be reduced by performing massively parallel custom syntheses on microchips[1,2]. This can now be achieved using a variety of methods, including ink-jet printing with standard reagents (Agilent), photolabile 5' protecting groups (Nimblegen/Affymetrix), photo-generated acid deprotection (Atactic/Xeotron) and electrolytic acid/base arrays (Oxamer/Combimatrix). However, current microchips have very low surface areas and hence only small amounts of oligonucleotides can be produced. When released into solution, the oligonucleotides are present at pictomolar or lower concentrations per sequence, concentrations that are insufficiently high to drive bimolecular priming reactions efficiently.

A potential solution to this scale problem would be to amplify the

oligonucleotides obtained from the microchips from roughly as little as 10^5 (or 10^9 for low density arrays) up to 10^9 (or 10^{12}) molecules of each sequence, thereby permitting subsequent selection and assembly steps. An overview of the integrated process is presented in Fig. 1.

For this amplification method (shown in Fig. 2a), oligonucleotides flanked by short, generic adaptor sequences are synthesized on a programmable microchip. This generates a pool of 10^2–10^5 different oligonucleotides, which can be released from the microchips by chemical or enzymatic treatment. Released oligonucleotides are amplified by polymerase chain reaction (PCR) using primers that contain type-IIS restriction enzyme recognition sites. Digestion of the PCR products with the corresponding restriction enzyme(s) yields sufficient amounts of unadulterated oligonucleotide sequences to be used for gene or genome assembly.

We first demonstrated the feasibility of this approach with Atactic/Xeotron 4K (that is, 3,968 synthesis chambers) photo-programmable microfluidic microarrays[1]. To monitor oligonucleotide synthesis and cleavage from the microchip, the $5'$ ends of the oligonucleotides were coupled with fluorescein. The microchip was scanned with a microarray scanner before and after cleavage, and the images are shown in Fig. 2b. The cleaved portions of the oligonucleotides were hybridized onto a 'quality-assessment (QA)-chip' synthesized with complementary oligonucleotide sequences (Fig. 2c). These results demonstrated that individual oligonucleotides were synthesized and nearly completely released from the microchip in quantities that can be measured by a QA-chip hybridization process. The typical yield of oligonucleotide released from each chamber of the 4K microchip is about 5 fmoles, as determined by quantitative PCR[1]. Using primers that anneal specifically to the generic adaptors flanking the oligonucleotide sequences, PCR reactions were carried out to amplify the oligonucleotides more than a million-fold.

Mutations incurred during oligonucleotide synthesis are a major source of errors in assembled DNA molecules, and are costly and difficult to eradicate[4,5]. We developed a simple, stringent hybridization-based method to remove oligonucleotides with such mutations. To select against mutations in gene construction oligonucleotides, these oligonucleotides were hybridized sequentially to two pools of bead-immobilized short complementary selection oligonucleotides that together span the entire length of the construction oligonucleotides (Fig. 3a). All selection oligonucleotides were designed to have nearly identical melting temperatures by varying their lengths. Under appropriate hybridization conditions, imperfect pairs between selection and construction oligonucleotides due to base-mismatch or deletion have lower melting temperatures and are unstable. After the cycles of hybridization, wash and elution, oligonucleotides with sequences that perfectly match the selection oligonucleotides are preferentially retained and enriched. Digestion of the PCR products with type-IIS restriction enzymes removed the generic primer sequences from both ends of the oligonucleotides (Fig. 3b, lanes 1–2). In these experiments the amplification tags are removed just before selection. However, if the digestion were deferred, the oligonucleotides could be re-amplified by PCR and subjected to further rounds of hybridization selection. Because the probability of complementary mutations occurring at matching positions on construction and selection oligonucleotides is miniscule, in principle most oligonucleotides with mutations can be eliminated by this selection procedure.

Like construction oligonucleotides, selection oligonucleotides were also synthesized and released from programmable microarrays. Selection oligonucleotides with arms were amplified by PCR, and the strands complementary to the gene construction oligonucleotide were labelled with biotin at the $5'$ end and selectively immobilized on streptavidin beads. The unlabelled strands were denatured and removed. As demonstrated in Fig. 3b (lane 3), immobilized selection oligonucleotides selectively retained the correct 50-bp construction oligonucleotides.

Such oligonucleotides are suitable for gene assembly. To facilitate automation, we developed a single-step polymerase assembly multiplexing (PAM) reaction for multiple gene syntheses from a single pool of oligonucleotides. Single-fragment assembly methods
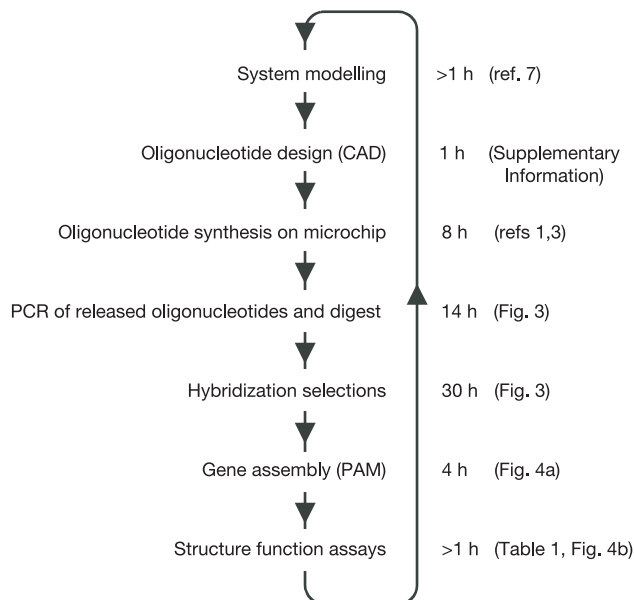

Figure 1 A flow chart for the design, synthesis and analysis of multiple genes in pools. Estimates of the current process timing (not always the minimum possible times) and sources of additional details are listed.

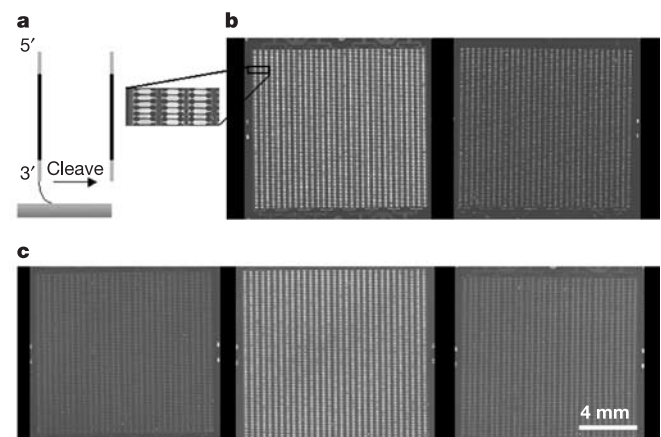| System modelling | >1 h | (ref. 7) |
Oligonucleotide design (CAD)	1 h	(Supplementary Information)
Oligonucleotide synthesis on microchip	8 h	(refs 1,3)
PCR of released oligonucleotides and digest	14 h	(Fig. 3)
Hybridization selections	30 h	(Fig. 3)
Gene assembly (PAM)	4 h	(Fig. 4a)
Structure function assays	>1 h	(Table 1, Fig. 4b)

Figure 2 Preparation of free oligonucleotides from a custom microarray. **a**, Diagram of synthesis and cleavage of a PCR-amplifiable oligonucleotide from a microchip surface. The portion of the oligonucleotide used for gene construction is in black; PCR primer-adaptors are shown in grey. **b**, Synthesis and cleavage of oligonucleotides from a Xeotron/Atactic 4K photo-programmable microfluidic microchip. Left: fluorescent scanning micrograph of an oligonucleotide array before cleavage. Insert: details of microfluidic chambers and connecting channels. Right: array after cleavage. **c**, Hybridization of released fluorescein (FAM)-labelled oligonucleotides to a QA-chip. Left: before hybridization; middle: after hybridization; right: after stripping of hybridized oligonucleotides.

Table 1 **Comparison of sequence errors generated by various methods**

Method	Total bp	Transition	Transversion	Deletion	Addition	bp per error
Hybridization selection (PAM)	23,641	7	3	5	2	1,394
PAGE selection (PAM)	24,546	28	12	11	3	455
No selection (PAM)	9,243	25	13	19	1	159
No selection (ligation)	6,093	6	6	22	4	160

We performed χ^2 tests for hybridization selection versus PAGE selection ($P = 2 \times 10^{-5}$), and hybridization selection versus no selection ($P = 2 \times 10^{-21}$). Only the constructs in the row labelled 'PAGE selection' involved gel purification.

have traditionally used two or three steps (ligation, assembly and PCR)[4–6]. For PAM, gene-flanking primer pairs were added to the pool of gene-construction oligonucleotides (with the primer pairs at a higher concentration than the oligonucleotides), together with thermostable polymerase and dNTPs. Extension of overlapping oligonucleotides and subsequent amplification of multiple full-length genes were accomplished in a closed-tube, one-step reaction using a thermal cycler. Different generic adaptor sequences can be incorporated into the ends of each gene or gene set, and a set of complementary adaptor-primer pairs can be pre-synthesized to avoid the cost of synthesizing gene-specific PAM primer pairs and to facilitate automation (for example, 96 or 384 generic adaptors to match standard multi-well plates).

To determine the efficiency of the hybridization-selection method to eliminate mismatch mutations[7], we constructed genes using the same pool of microchip-synthesized oligonucleotides purified in three different ways: unpurified, polyacrylamide gel electrophoresis (PAGE)-purified or hybridization-purified. These genes were cloned and random clones from each category were sequenced in both directions to determine error types and rates for each category. As shown in Table 1, genes synthesized with unpurified oligonucleotides have the highest error rates (1 in 160 bp); the method of gene assembly (using ligation or PAM) made little difference. PAGE purification of oligonucleotides reduced the error rate to 1 in 450 bp, mainly through removal of deletion mutations. This rate is comparable to figures reported by other groups using PAGE purification[4,5]. With hybridization selection, the error rate was further reduced to approximately 1 in 1,394 bp.

As an example of the usefulness of this technology for large-scale synthetic biology projects[8], we used a microchip to redesign and synthesize codon-altered versions of the 21 protein-encoding genes that constitute the *E. coli* small ribosomal subunit. We had previously noticed while using the natural versions of these 21 proteins[9] that translational efficiencies were very low *in vitro*, even though *in vivo* the proteins have high expression levels. Redesigning codon usage is a way to increase protein translation efficiencies, although it is more challenging to accomplish when starting with nearly ideal codons. Because many other proteins are expressed well in this *in vitro* system, we hypothesized that some of the problem was due to secondary structure (possibly exacerbated by the fact that the rate of T7 polymerase-mediated transcription is eightfold higher than translation[10,11]). We attempted to replace codons with sequences likely to have less secondary structure (for example, by lowering G + C content). Our computer-aided design software (CAD-PAM) designed overlapping 50-bp oligonucleotide sequences (embedded in 70-mers) for the 21 ribosomal genes and synthesized them all on a 4K Xeochip. These oligonucleotides were processed and hybridization-selected with selection oligonucleotides, and were then used to construct the 21 ribosomal genes in multiple PAM reactions (Fig. 4a). Error-free clones were tested in *E. coli* using coupled *in vitro* transcription–translation reactions. The translation profiles of the synthetic genes are shown in Fig. 4b. A number of codon-altered genes had higher translation levels in the *E. coli* extract compared with their respective wild-type genes. We next combined these 21 genes using sequential PAM reactions to give a pool of ~14.6 kb assemblies (Fig. 4c) by introducing unique ~30-mer overlapping linkers between gene units and performing sequential PAM reactions. Correct assembly was confirmed by sequencing on average four individual clones from every over-
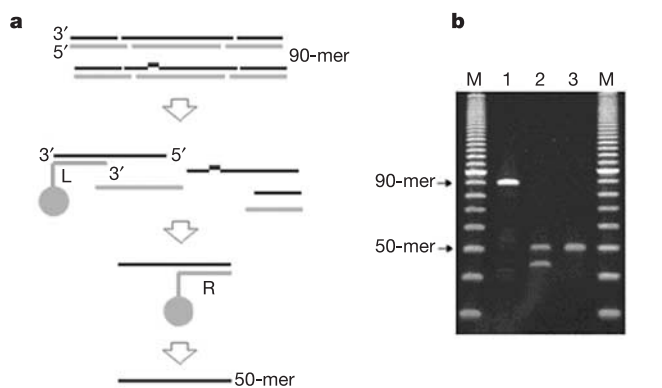

Figure 3 Hybridization selection of microchip-synthesized oligonucleotides. **a**, Diagram of the hybridization selection strategy. 90-mer oligonucleotides (upper strands black, lower strands grey) are cut with type-IIS restriction enzymes to release hybrids of 50-mers and complementary 44-mers, some of which have incorrect sequences (indicated by a bulge in the upper strand of the second 90-mer oligonucleotide). Only the correct upper 50-mer strand hybridizes well with left (L) then right (R) selection oligonucleotides (immobilized on beads in grey). **b**, A denaturing PAGE gel separating PCR-amplified oligonucleotides (90-mer, lane 1), *Bsa*I/*Bse*RI-digested oligonucleotides (50- and 44-mer, lane 2) and hybridization-selected oligonucleotides (50-mer, lane 3). Ten-base-pair markers are shown in lanes marked M.
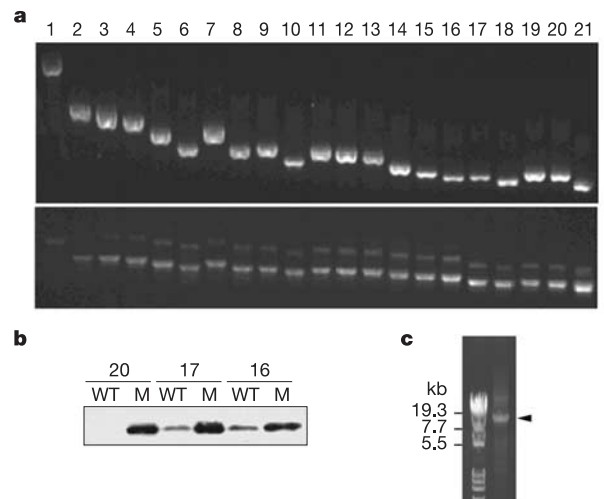

Figure 4 Synthetic gene constructs and their translation products. **a**, Twenty-one engineered ribosomal protein genes constructed using PAM reactions (upper gel) and their *in vitro* transcribed RNAs (lower set of bands in the lower gel). **b**, Ribosomal proteins translated from their synthetic genes. Western blot comparison of translation efficiency between wild-type (WT) and engineered (M) versions of three ribosomal protein genes. **c**, A 14.6-kb operon of 21 ribosomal genes assembled using sequential PAM reactions. Left lane, DNA molecular weight marker IV (Roche).

lapping DNA segment generated by high-fidelity PCR reactions that together covered the whole construct. By starting with correct input gene sequences, and through repeated high-fidelity, polymerase-based extension reactions, the assembly process resulted in a lower error rate (about 1 in 7,300 bp) than any of the methods shown in Table 1 (all of which started with oligonucleotides containing synthetic errors). This clearly demonstrated that a major source of error for gene assembly comes from oligonucleotide chemical synthesis rather than polymerase proofreading activity. Although the increasing length of the PCR products might be expected to reduce yield in the later assemblies, the number of reaction components decreases and so the efficiency remains high. Eventually when PAM length does become limiting, homologous recombination might allow assembly in the megabase range.

In this demonstration, a small fraction of a microchip's synthetic capacity was used. Massively parallel microchip-based oligonucleotide synthesis might increase yields in oligonucleotide synthesis from 9 bp per US$ to 20 kb per US$, depending on the type of microchip used and the number of oligonucleotides on a microchip. To achieve few or no errors in megabase-scale assemblies, the progress here on mismatch errors must be extended. In addition to the physical methods described here, we and others are using protein-based methods (for example, using MutS[12] or MutHLS mismatch correction proteins) already shown to be capable of exceeding by tenfold the fidelity of DNA polymerases (the latter being 6×10^{-6} errors per base per cycle)[13]. This could significantly reduce the money spent on sequencing and potentially eliminate the need for error correction using directed-mutagenesis methods. Ultra-low-cost DNA sequencing[14] could also have a large impact if properly integrated. The next stage in testing the limits to simultaneous synthesis and assembly will employ 95,000–382,000 oligonucleotides per $700 microchip from Nimblegen (yielding 2–18 Mbp). The first such assembly has been successful (see http://arep.med.harvard.edu/SBP/). Overcoming the challenge of repeated sequences in desired synthetic constructs may require approaches like hierarchical assembly, assembly focused on only the unique ends, and/or multiplex size selections of assembly products. Using geometrical constraints to force a certain order of assembly illustrates another potential strength of microfluidic syntheses. Improvement of error rates enables a variety of applications without cloning or sequencing.

Although oligonucleotide assembly is an established procedure[6,15,16], our procedures for elution from DNA microchips, oligonucleotide amplification, mismatch error correction, multiplexing and microfluidic integration are new. Accurate, low-cost multiplex gene syntheses will be a powerful tool for synthetic biology[8] and complex nanostructures in general[17]. For example, this technology is currently enabling us to improve and test components needed for the synthesis of ribosomes *in vitro*. The rapid prototyping of individual genes on short linear templates has greatly aided the debugging of this complex system. Our ability to remap a whole set of genes from the *Mycoplasma*[18] codons to those of *E. coli* (for example, by eliminating the UGA 'stop' codon and changing G + C content from 25% to 51%) will help us to calibrate proteomics experiments, test *de novo* protein designs and identify new biochemical activities, including those that are missing from our list of the core components of novel self-replicating systems[8]. □

Methods

Design of sequences

Gene and oligonucleotide sequences were designed using the Java program CAD-PAM, to be described in detail elsewhere (J.T., H.G. and G.C., manuscript in preparation). Basically, CAD-PAM uses constraints on the amino acid sequences, codon usage, messenger RNA secondary structure and restriction enzymes used to release the construction oligonucleotides in order to create nearly optimal, overlapping sets of *n*-mer (typically 50-mer) construction oligomers and shorter selection oligomers (typically 26-mer). The melting temperatures (T_m) of overlapping regions between adjacent gene construction

oligonucleotides or between construction and selection oligonucleotides were equalized. The selection oligonucleotides were padded with extra adenine residues to keep oligomer length constant (70-mers) for optional size selection (not used for typical PAM). T_m values were calculated using the nearest neighbour method[19]. Codons can be fixed or altered to allow expression improvements. The sequences of the oligonucleotides used in this study are provided in Supplementary Table 1.

Microchip synthesis, amplification and selection of oligonucleotides

Oligonucleotides were synthesized on photo-programmable microfluidic microchips with a phosphate at the 5′ end and the 3′ end coupling to the 3′-hydroxy terminus of a uracil residue (X. Gao, unpublished work). After synthesis, the oligonucleotides were cleaved either with RNase A or by ammonium hydroxide treatment (used for deprotection as in standard oligonucleotide syntheses) followed by precipitation. Gene construction oligonucleotides that had been PCR amplified with 20-mers (initially complementary to the terminal ten bases) were digested with the type-IIS restriction enzymes *Bsa*I and *Bse*RI (without gel purification except for the 'PAGE' controls in Table 1). While this paper was in review a related method using the enzyme *Mly*I appeared[20]. Immobilization of biotin-labelled selection oligonucleotides on magnetic streptavidin beads (Dynal) and removal of the non-biotinylated strand were done as described[21]. Construction oligonucleotides were denatured at 95 °C for 3 min and hybridized to selection oligonucleotides in hybridization buffer (5 × SSPET buffer, 50% formamide, 0.2 mg ml^{-1} BSA) for 14–16 h at 42 °C on a rotor. Beads were washed three times with 0.5 × SSPET and three times with wash buffer (20 mM Tris-HCl pH 7.0, 5 mM EDTA, 4 mM NaCl) at room temperature. The construction oligonucleotides were recovered by denaturation in 0.1 M NaOH for 15 min and subsequent neutralization.

Polymerase assembly multiplexing reactions

PAM reactions were carried out in 25-μl reactions containing 2 μl of oligonucleotide mixtures, 0.4 μM of each of the gene-end primer pairs, 1 × dNTP mixture and 0.5 μl of Advantage 2 polymerase mixture in 1 × buffer (Clontech Advantage 2 PCR kit). Samples were denatured at 95 °C for 3 min, then underwent 40–45 thermal cycles of 95 °C for 30 s, 49 °C for 1 min and 68 °C for 1 min kb^{-1}, then finished at 68 °C for 10 min. Sequential PAM reactions were used to combine multiple genes. First, His6-tagged linear expression constructs of the correct sequences of 21 ribosomal protein genes were pre-constructed by PCR using an RTS *E. coli* linear template generation kit (Roche). These constructs were then used as templates in separate PCR reactions where unique ~30-mer linkers with identical T_m (0.4 μM of each, Integrated DNA Technologies, Inc.) were introduced to create enough overlapping sequences between genes for secondary PAM reactions (see Supplementary Table 2). In these, three large fragments were made in separate Roche Expand long template PCR reactions: RS1–5 (1–5,513), RS6–13 (5,483–10,526) and RS14–21 (10,497–14,593). These fragments were gel-purified and assembled into a full 14,593-bp operon in the final assembly reaction using RS1–21 (1–14,593). For the last two assemblies, samples were denatured at 92 °C for 2 min, followed by 10 thermal cycles at 92 °C for 30 s, 65 °C for 1 min and 68 °C for 1 min kb^{-1}, then followed by 25 additional cycles at 92 °C for 30 s, 65 °C for 1 min, and 68 °C for 1 min kb^{-1} plus 10 s per cycle, and finished at 68 °C for 10 min.

Coupled *in vitro* transcription and translation

Assembled genes were cloned and error-free clones were selected by sequencing. Linear constructs for *in vitro* protein expression were made using Roche RTS *E. coli* linear template generation set, His-tag. *In-vitro*-coupled transcription and translation was performed using a Roche Rapid Translation System RTS 100 *E. coli* HY kit. Proteins were detected by western blotting with an anti-His6-peroxidase antibody (Roche) using standard procedures.

1. Zhou, X. *et al.* Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.* **32,** 5409–5417 (2004).
2. Fodor, S. P. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251,** 767–773 (1991).
3. Gao, X. *et al.* Flexible DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res.* **29,** 4744–4750 (2001).
4. Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297,** 1016–1018 (2002).
5. Smith, H. O., Hutchison, C. A. III, Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA* **100,** 15440–15445 (2003).
6. Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164,** 49–53 (1995).
7. Eason, R. G. *et al.* Characterization of synthetic DNA bar codes in *Saccharomyces cerevisiae* gene-deletion strains. *Proc. Natl Acad. Sci. USA* **101,** 11046–11051 (2004).
8. Forster, A. C. & Church, G. M. A synthetic biology project. *Nature* (submitted).
9. Culver, G. M. & Noller, H. F. Efficient reconstitution of functional *Escherichia coli* 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5,** 832–843 (1999).
10. Iost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes *in vivo. J. Bacteriol.* **174,** 619–622 (1992).
11. Iost, I. & Dreyfus, M. mRNAs can be stabilized by DEAD-box proteins. *Nature* **372,** 193–196 (1994).
12. Carr, P. A., Park, J. S., Lee, Y.-J., Yu, T., Zhang, S. & Jacobson, J. M. Protein-mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res.* (in the press).
13. Smith, J. & Modrich, P. Removal of polymerase-produced mutant sequences from PCR products. *Proc. Natl Acad. Sci. USA* **94,** 6847–6850 (1997).

14. Shendure, J., Mitra, R., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nature Rev. Genet.* **5**, 335–344 (2004).

15. Mullis, K. *et al.* Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **LI**, 263–273 (1986).

16. Dillon, P. J. & Rosen, C. A. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *BioTechniques* **9**, 298–300 (1990).

17. Liu, D., Park, S. H., Reif, J. H. & LaBean, T. H. DNA nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. *Proc. Natl Acad. Sci. USA* **101**, 717–722 (2004).

18. Jaffe, J. D., Berg, H. C. & Church, G. M. Proteogenomic mapping reveals genomic structure and novel proteins undetected by computational algorithms. *Proteomics* **4**, 59–77 (2004).

19. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA* **83**, 3746–3750 (1986).

20. Richmond, K. E. *et al.* Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32**, 5011–5018 (2004).

21. Espelund, M., Stacy, R. A. & Jakobsen, K. S. A simple method for generating single-stranded DNA probes labeled to high activities. *Nucleic Acids Res.* **18**, 6157–6158 (1990).

∙∙∙

The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes

Joseph T. Wade, Daniel B. Hall & Kevin Struhl

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

∙∙∙

Ribosomal protein (RP) genes in eukaryotes are coordinately regulated in response to growth stimuli and environmental stress, thereby permitting cells to adjust ribosome number and overall protein synthetic capacity to physiological conditions[1–5]. Approximately 50% of RNA polymerase II transcription is devoted to RP genes[5]. The transcriptional regulator Rap1 binds most yeast RP promoters[6], and Rap1 sites are important for coordinate regulation of RP genes[7–10]. However, Rap1 is not the specific regulator that controls RP transcription because it also functions as a repressor, and many Rap1-activated promoters are not coordinately regulated with RP promoters[11,12]. Here we show that the transcription factors Fhl1 and Ifh1 associate almost exclusively with RP promoters; association depends on Rap1 and (to a lesser extent) a DNA element at many RP promoters. Ifh1 is recruited to promoters via the forkhead-associated (FHA) domain of Fhl1; the level of Ifh1 associated with RP promoters determines the level of transcription; and environmental stress causes a marked reduction in the association of Ifh1, but not Fhl1 or Rap1. Thus, Ifh1 association with promoters is the key regulatory step for coordinate expression of RP genes.

Yeast cells contain ~150 copies of the ribosomal DNA locus and 137 RP genes that encode one or two copies of each of 78 proteins. Transcription of the RP genes is coordinately regulated in accord with the cellular growth rate in a manner that requires protein kinase A and the TOR pathway[8,13–15]. In addition, RP transcription is rapidly and coordinately downregulated in response to a variety of environmental insults such as heat shock, amino-acid starvation and osmotic shock. Coordinate regulation of RP genes has a major biological impact on the overall protein synthetic capacity and growth of the cell, and 50% of RNA polymerase II transcription in yeast is devoted to RP genes[5].

Despite the biological importance of coordinate regulation of RP genes, information about the mechanism of this regulation is rather incomplete. Approximately 90% of RP promoters contain predicted Rap1 binding sites[5,16], and Rap1 is bound to essentially all such RP promoters *in vivo*[6]. Rap1 sites are important for growth-regulated expression of natural RP promoters[7–9], and a 41-base-pair (bp) fragment containing two Rap1 sites from an RP promoter is sufficient to mediate growth regulation[10]. However, Rap1 also binds and activates many non-RP promoters that are not regulated in the same manner as RP promoters. Furthermore, the Rap1-dependent activator that regulates transcription of RP genes recruits TFIID, whereas the Rap1-dependent activator that regulates transcription of glycolytic genes does not[10]. Taken together, these results strongly suggest the existence of an unknown protein(s) that is specifically involved in the coordinate regulation of RP genes.

An extensive genome-wide analysis of target sites for over 100 yeast DNA-binding transcription factors revealed Fhl1, a protein containing a fork head DNA-binding domain[17], as binding specifically to RP promoters[18]. We confirmed this result by performing chromatin immunoprecipitation (ChIP) on Fhl1 coupled with analysis on microarrays containing essentially all yeast promoter regions (Fig. 1a; Supplementary Table 1). Using a stringent cutoff of fivefold enrichment, we identified 79 targets, 76 of which are RP promoters. At this level of stringency, Fhl1 is almost exclusively specific to RP promoters. It is highly likely that Fhl1 interacts with additional RP promoters, as 80% (103 out of 129) of detectable RP promoter regions are found in the top 5% of the Fhl1-bound targets (Fig. 1a).

We also determined the genome-wide association of Ifh1, a protein that interacts genetically with Fhl1 (ref. 19). The binding profile of Ifh1 is strikingly similar to that of Fhl1, located almost exclusively at RP promoters (Fig. 1a; Supplementary Table 1). Of the 56 targets showing at least fivefold enrichment, 54 were RP promoters, and it is likely that Ifh1 associates with additional RP promoters (Fig. 1a). Remarkably, 94% (51 out of 54) of Ifh1 targets defined in this manner are also Fhl1 targets. Fhl1 and Ifh1 do not bind to non-RP, Rap1-containing promoters or to the ribosomal DNA loci (Fig. 1a, b; data not shown). Ifh1 and Fhl1 do not associate with eight of the nine RP promoters that do not bind Rap1 *in vivo*[6,16]. Thus, Fhl1 and Ifh1 bind almost exclusively to a common set, but not all RP promoters, and binding of these proteins seems to be influenced by Rap1.

Conserved DNA sequence motifs among Fhl1- and Ifh1-bound promoters include the Rap1 binding site, an A-rich stretch implicated in RP transcription[20], and a close match to a sequence element, 'Motif 213', recently identified in a subset of RP promoters by a computational analysis designed to predict gene expression using DNA sequence information[21] (Fig. 2a; Supplementary Fig. 1). Analysis of RP promoters that do not display fivefold enrichment of Fhl1 and Ifh1 identifies the Rap1 site and the A-rich stretch, but not the third motif. We therefore name the third motif IFHL, and a search of this motif across the yeast genome shows that it is highly over-represented at RP gene promoters ($P < 0.001$). Although the presence of the IFHL motif correlates with binding of Ifh1 and Fhl1 *in vivo*, there is not a strict one-to-one correspondence because 38 of the Fhl1 targets do not contain a clear IFHL motif.

Mapping of Fhl1 and Rap1 binding sites across three RP promoters indicate promoter-specific differences in the relative locations of these proteins (Fig. 2b–d). At *RPL12A*, Rap1 and Fhl1 associate with discrete regions, with Rap1 binding near the two predicted Rap1 sites and Fhl1 binding near the IFHL motif (Fig. 2b). In contrast, Fhl1 binding at *RPS11B* and *RPL40A* matches that of